Mode d'emploi

On passera suffisamment de temps sur le **I** et le **II** pour bien comprendre la notion, essentielle en probabilités, de conditionnement. Il est donc conseillé de faire tous les exercices. Les exercices d'application typique de la formule de Bayes ne seront pas beaucoup revus en spé...et quand on en a fait deux ou trois, on voit bien comment cela fonctionne.
La suite (variables aléatoires) est à lire attentivement plusieurs fois, mais elle ne doit pas être trop difficile si on a bien compris **I** et **II**. Il y a d'ailleurs beaucoup moins d'exercices (les vacances sont quand même des vacances), autant les faire consciencieusement.

P1 : Espaces probabilisés finis

I Probabilités sur un ensemble fini

I.1 Univers, événements, probabilité

Considérons une « expérience aléatoire » typique : le lancer d'un dé (aléa et hasard viennent des mots latin et arabe signifiant « dé »...même le mot chance vient du jeu d'osselets puis de dés, voir par exemple le *Robert historique de la langue française*). On s'intéresse bien sûr au nombre de points figurant sur la face supérieure du dé. Les résultats possibles sont donc

$$1,2,3,4,5,6$$

Définition On appelle univers l'ensemble des résultats (ou issues, ou réalisations...) d'une expérience aléatoire.

L'univers associé à l'expérience aléatoire précédente (jet de dé) est donc naturellement l'ensemble $\{1,2,3,4,5,6\}$.

Supposons le dé équilibré. Par raison de symétrie, on aura autant de chance d'obtenir un 2 qu'un 6. On dira que toutes les issues sont équiprobables. On pourrait alors définir une probabilité sur l'univers $\Omega = \{1,2,3,4,5,6\}$ par

$$\forall i \in \{1,\dots,6\} \qquad p(i) = p_i = \frac{1}{6}$$

Evidemment, il se peut que le dé soit « pipé ». Pour définir une probabilité convenable, la raison de symétrie ne fonctionnera plus, on aura recours aux statistiques…
Supposons que l'on s'intéresse à la parité du résultat (il vaudrait mieux dans ce cas jouer à pile ou face, mais ça viendra). On considère l'événement aléatoire « le résultat est pair ». On lui associe la partie

$$A = \{2, 4, 6\}$$

de Ω. Et on définit naturellement :

Définition Soit Ω un univers fini. On appelle événement toute partie de Ω ; on appelle événement élémentaire tout singleton $\{\omega\}$, où $\omega \in \Omega$.

Revenons à notre probabilité ; on associe à l'événement A la probabilité

$$p(A) = \sum_{i \in A} p(i) = 3 \times \frac{1}{6} = \frac{1}{2}$$

Plus généralement, on pourrait définir, sur un ensemble fini Ω, une probabilité p comme étant une application $p : \Omega \to [0, 1]$ telle que

$$\sum_{\omega \in \Omega} p(\omega) = 1$$

puis définir la probabilité d'un événement A par

$$p(A) = \sum_{\omega \in A} p(\omega)$$

mais cette définition serait hétérogène : p serait définie à la fois sur Ω et sur $\mathscr{P}(\Omega)$. Il vaut mieux donc définir p seulement sur $\mathscr{P}(\Omega)$, on pourra donc écrire $p(\{\omega\})$, mais pas $p(\omega)$.

Définition On appelle probabilité sur un ensemble fini Ω toute application P de $\mathscr{P}(\Omega)$ dans $[0,1]$ telle que $P(\Omega) = 1$ et telle que

$$\forall (A,B) \in \mathscr{P}(\Omega)^2 \qquad A \cap B = \emptyset \;\Rightarrow\; P(A \cup B) = P(A) + P(B)$$

Un espace probabilisé fini est un couple (Ω, P) où Ω est un ensemble fini et P une probabilité sur Ω

Attention au vocabulaire : une probabilité sur Ω n'est pas définie sur Ω, mais sur $\mathscr{P}(\Omega)$

Dans la suite, si (Ω, P) est un espace probabilisé fini, on se permettra d'appeler événement toute partie de Ω.

I.2 Quelques univers

a. Poker

On considère l'expérience aléatoire suivante : tirer 5 cartes dans un jeu de 52 cartes. Définir un univers associé à cette expérience. Quel est son cardinal ?

b. Bridge

On considère l'expérience aléatoire suivante : tirer 13 cartes dans un jeu de 52 cartes. Définir un univers associé à cette expérience. Quel est son cardinal ?

c. Tirages répétés

Quel univers correspond à un tirage Pile ou Face répété n fois ?

d. Des objets dans des boîtes

On considère l'expérience suivante : placer au hasard r objets $x_1, \ldots, x_r$ dans n boîtes $b_1, \ldots, b_n$. Définir un univers associé à cette expérience. Quel est son cardinal ?

e. Des objets indiscernables dans des boîtes

On considère l'expérience suivante : placer au hasard r objets indiscernables dans n boîtes $b_1, \ldots, b_n$. Définir un univers associé à cette expérience. Quel est son cardinal ?
On comprendra bien la différence avec la question précédente, dans laquelle, par exemple, tous les objets sauf x_1 dans b_1 et x_1 dans b_2 d'une part, tous les objets sauf x_2 dans b_1 et x_2 dans b_2 d'autre part sont deux issues (ou réalisation) différentes. Alors qu'ici, tous les objets sauf un dans b_1 et le dernier dans b_2, c'est une issue.

f. Des objets indiscernables dans des boîtes indiscernables

On considère l'expérience suivante : placer au hasard r objets indiscernables dans n boîtes indiscernables. Définir un univers associé à cette expérience.

I.3 Première propriété : additivité

Proposition Soit P une probabilité. Si $A_1, \ldots, A_m$ sont des événements deux à deux disjoints ($m \geq 2$), alors

$$P\left(\bigcup_{i=1}^{m} A_i\right) = \sum_{i=1}^{m} P(A_i)$$

(Démonstration par récurrence sur m). On appelle cette propriété « additivité ». Il arrive d'ailleurs que la réunion **disjointe** $\bigcup_{i=1}^{m} A_i$ soit notée $\sum_{i=1}^{m} A_i$ (insistons : il n'est cohérent de noter une réunion comme ceci que lorsqu'elle est disjointe : $(i \neq j) \Rightarrow A_i \cap A_j = \emptyset$). Cette notation est aussi cohérente avec les fonctions caractéristiques, que l'on verra plus tard.

Corollaire Si P est une probabilité sur Ω, si $A = \{\omega_1, \ldots, \omega_m\}$ est un événement, alors

$$P(A) = \sum_{i=1}^{m} P(\omega_i)$$

Autrement dit, une probabilité est déterminée par les probabilités des singletons.

I.4 Probabilité uniforme

Définition Soit $N = \text{Card}(\Omega)$. La probabilité uniforme sur Ω est l'unique probabilité telle que

$$\forall \omega \in \Omega \qquad P(\{\omega\}) = \frac{1}{N}$$

On a alors, pour tout événement A,

$$P(A) = \frac{\text{Card}\, A}{\text{Card}\, \Omega}$$

Dans le cas de la probabilité uniforme sur un univers fini, le calcul de la probabilité d'un événement équivaut au calcul de son cardinal. C'est donc un problème de dénombrement.

Exercice : On lance deux fois un dé « équilibré ». Quelle est la probabilité pour que la somme des deux résultats obtenus soit égale à 2 ? égale à 7 ? supérieure ou égale à 8 ?

Exercice : On tire 5 cartes d'un jeu de 32 cartes. Quelle est la probabilité qu'il y ait parmi ces 5 cartes au moins un Roi et au moins un Coeur ? (si on a le Roi de Coeur, on considère qu'on a bien un Roi et un Coeur).

I.5 Autres propriétés

Définition-Proposition Soit (Ω, P) un espace probabilisé. Soit A un événement (c'est-à-dire une partie de Ω). L'événement $A^c = \Omega \setminus A$ (complémentaire de A) est appelé « événement contraire de A ». Sa probabilité est

$$P(A^c) = 1 - P(A)$$

Démonstration : On a $A \cap A^c = \emptyset$, et $A \cup A^c = \Omega$, on peut donc écrire

$$P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$$

Définition-Proposition Soit (Ω, P) un espace probabilisé, A et B deux événements. Les événements $A \cup B$ et $A \cap B$ sont respectivement appelés « A ou B » et « A et B ». On a

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

On peut faire un petit dessin. . .et considérer par exemple le complémentaire B' dans B de $A \cap B$: l'additivité donne alors

$$P(A \cap B) + P(B') = P(B)$$

Mais on a aussi $A \cup B = A \cup B'$ et $A \cap B' = \emptyset$ ce qui redonne encore par additivité :

$$P(A \cup B) = P(A) + P(B')$$

et le résultat s'en déduit.

Proposition Si A et B sont deux événements d'un espace probabilisé (Ω, P), on a

$$A \subset B \quad \Rightarrow P(A) \leq P(B)$$

Cette propriété est la croissance, et se déduit de l'existence, lorsque $A \subset B$, d'une partie C telle que

$$A \cap C = \emptyset \qquad \text{et} \qquad A \cup C = B$$

(autrement dit, $C = B \setminus A$) ce qui donne

$$P(A) + P(C) = P(B)$$

et permet de conclure car $P(C) \geq 0$.
On dit parfois « A implique B » pour $A \subset B$.

Définitions L'événement impossible est $\emptyset$. Sa probabilité est 0 (on a en effet $\emptyset \cup \emptyset = \emptyset$ et $\emptyset \cap \emptyset = \emptyset$, on peut alors appliquer l'additivité). Deux événements A et B sont dits incompatibles lorsque $A \cap B = \emptyset$, autrement dit lorsque « A et B » est l'événement impossible.

Dans un univers fini, la distinction entre « événement impossible » et « événement de probabilité nulle » n'a pas grand intérêt (si des événements élémentaires sont de probabilité nulle, on peut sans dommage les enlever à l'univers). En revanche, il faut bien faire attention à ne pas confondre ces deux notions dans d'autres cas : si on tire au hasard un nombre dans le segment $[0,1]$*, avec probabilité uniforme, la probabilité que ce nombre soit dans* $[a,b]$ *où* $a \leq b$ *est* $b-a$ *(rapport de la mesure du segment* $[a,b]$ *au segment* $[0,1]$*). Mais la probabilité que ce nombre soit égal à* α *est nulle pour tout* α*. Or cet événement n'est pas impossible.*

Exercice (paradoxe de Monty Hall) : On consultera avec plaisir et intérêt la longue histoire des controverses autour de ce paradoxe, par exemple sur la page wikipedia qui lui est consacrée. Donnons-en une version brève.
Un candidat participe à un jeu télévisé, dans lequel il fait face à trois portes identiques, fermées. Derrière deux de ces portes, il n'y a rien. Derrière la troisième, un fabuleux trésor, qui lui reviendra s'il choisit cette porte.
Le candidat choisit donc une porte. Puis intervient l'animateur du jeu, qui sait ce qu'il y a derrière chacune des portes. Il est donc en mesure de dire au candidat, en désignant une des deux portes non choisies par celui-ci : derrière cette porte, il n'y a rien. Le candidat a alors le choix : garder la porte qu'il avait initialement choisi, ou changer son choix (et donc, évidemment, choisir celle des deux autres portes que l'animateur n'a pas désignée). Que doit-il faire ?

Exercice (tirage dans une urne avec remise ; loi binomiale)

1. Dans une urne, il y a n_1 boules blanches, n_2 boules rouges. On effectue r tirages successifs avec remise (on tire une boule au hasard, on regarde sa couleur, on la remet dans l'urne. On fait cela r fois). Calculer la probabilité pour que l'on tire k fois une boule rouge ($0 \leq k \leq r$). On pourra noter $p = \dfrac{n_1}{n_1+n_2}$ et $q = 1-p = \dfrac{n_2}{n_1+n_2}$.
   On utilisera sans forcément l'expliciter l'indépendance des tirages : par exemple, P(la première boule est rouge et la deuxième est blanche)=P(la première boule est rouge)×P(la deuxième boule est blanche).
2. Cette fois, il y a dans l'urne n_1 boules blanches, n_2 boules rouges, n_3 boules noires. Calculer la probabilité pour qu'après m tirages avec remise on ait tiré exactement k boules blanches et ℓ boules rouges (et donc $m-k-\ell$ boules noires : $k+\ell \leq m$ nécessairement). On pourra noter
$$p = \frac{n_1}{n_1+n_2+n_3}\ ,\ q = \frac{n_2}{n_1+n_2+n_3}\ ,\ r = \frac{n_3}{n_1+n_2+n_3}$$

Exercice (tirage dans une urne sans remise ; loi hypergéométrique)

1. Dans une urne, il y a N_1 boules blanches, N_2 boules rouges. On effectue r tirages successifs sans remise (donc, nécessairement, $r \leq N = N_1+N_2$). Calculer la probabilité pour que l'on tire n_1 fois une boule blanche et n_2 fois une boule rouge (où $n_1 \leq N_1$, $n_2 \leq N_2$, $n_1+n_2 = r$).

2. Dans une urne, il y a N_1 boules blanches, N_2 boules rouges. On tire r boules simultanément (donc, nécessairement, $r \leq N = N_1 + N_2$). Calculer la probabilité pour que l'on tire n_1 fois une boule blanche et n_2 fois une boule rouge (où $n_1 \leq N_1$, $n_2 \leq N_2$, $n_1 + n_2 = r$).

II Conditionnement, indépendance

II.1 Conditionnement

Revenons sur le lancer de dé. On prend un dé, on effectue 1000 lancers et on obtient le bilan suivant (dans la seconde colonne, le nombre de fois où le chiffre indiqué dans la première a été obtenu) :

1	167
2	165
3	163
4	170
5	163
6	172

Deux questions intéressantes se posent : est-on prêt à considérer que le dé est « équilibré » ? sinon, quelle valeur attribuerait-on aux différentes probabilités d'obtenir 1,...,6 pour modéliser le lancer de ce dé ? une réponse mathématiquement fondée n'est pas si facile, mais elle se baserait sur l'étude des fréquences empiriques d'obtention des différents chiffres : 0,167 ; 0,165 ; 0,163 ; 0,170 ; 0,163 ; 0,172. La notion de fréquence empirique s'étend aux événements : ci-dessus, la fréquence empirique observée de l'événement « le chiffre obtenu est pair » est

$$\frac{165+170+172}{1000} = 0,507$$

Notons B cet événement. Et soit A l'événement « le chiffre obtenu est supérieur ou égal à 4 ». Le rapport

$$\frac{\text{nombre de fois où } A \text{ et } B \text{ sont réalisés}}{\text{nombre de fois où } B \text{ est réalisé}}$$

est aussi le quotient des fréquences empiriques

$$\frac{(170+172)/1000}{(165+170+172)/1000} \approx 0,675$$

s'interprète naturellement : si on sait que le résultat est pair, il y a environ deux chances sur trois pour qu'il soit supérieur ou égal à 4 (si le dé est équilibré). Cette fréquence empirique est intuitivement une approximation de la probabilité que A soit réalisé sachant que B l'est. On adopte donc la définition suivante :

Définition Soit (Ω, P) un espace probabilisé fini, B un événement tel que $P(B) > 0$. Pour tout événement A (i.e. pour toute partie A de Ω), on définit la probabilité conditionnelle de A sachant B par

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}$$

(*Se lit en général « P de A sachant B »*)

Proposition P_B est une probabilité sur Ω.

Démonstration P_B est une application définie sur $\mathscr{P}(\Omega)$, à valeurs réelles positives. Et même à valeurs dans $[0,1]$ par croissance de P. Clairement $P_B(\Omega) = 1$. Et, si $A \cap A' = \emptyset$, $(A \cap B) \cap \left(A' \cap B\right) = \emptyset$, et donc

$$\begin{aligned} P_B(A \cup A') &= \frac{P\left(\left(A \cup A'\right) \cap B\right)}{P(B)} \\ &= \frac{P\left((A \cap B) \cup \left(A' \cap B\right)\right)}{P(B)} \\ &= \frac{P\left(A \cap B\right) + P\left(A' \cap B\right)}{P(B)} \\ &= P_B(A) + P_B(A') \end{aligned}$$

La notation $P_B(A)$ est bien souvent préférable à la notation $P(A|B)$. Néanmoins, cette dernière est au programme, il faut la connaître.

II.2 Probabilités composées, probabilités totales, probabilités des causes

Proposition (**Formule des probabilités composées**)

Si A et B sont deux événements, si $P(B) \neq 0$, alors

$$P(A \cap B) = P(B)\ P_B(A)$$

Plus généralement, si $A_1, \ldots, A_m$ sont des événements, et si $P(A_1 \cap \ldots \cap A_{m-1}) \neq 0$, alors

$$P(A_1 \cap A_2 \ldots \cap A_m) = P(A_1) P_{A_1}(A_2) \ldots P_{A_1 \cap A_2 \cap \cdots \cap A_{m-1}}(A_m)$$

ou encore

$$P\left(\bigcap_{k=1}^{m} A_k\right) = \prod_{k=1}^{m} P_{\bigcap_{j=1}^{k-1} A_j}(A_k)$$

(avec la convention habituelle $\bigcap_{j=1}^{0} A_j = \Omega$, Ω étant neutre pour $\cap$. On remarque que $P_\Omega = P$).

Démonstration par récurrence.

Utilisation Lorsqu'un événement s'interprète naturellement comme intersection d'événements en général « successifs ».

Exercice : Une urne contient n boules blanches, n boules noires. On retire les boules une par une, au hasard. Quelle est la probabilité de tirer 1 boule blanche, puis une noire, puis une blanche, etc… ?

Exercice : On fait l'hypothèse (certes discutable) que la probabilité de naître un 2 décembre est la même que celle de naître un 18 juin, ou un 14 juillet…et on supposera en revanche que personne ne naît un 29 février.

1. Trouver la probabilité pour que, dans un groupe de dix personnes, il n'y en ait pas deux dont l'anniversaire tombe le même jour.

2. Même question pour un groupe de 40 personnes.
3. Avec une calculatrice ou un ordinateur, trouver le plus petit nombre N vérifiant : dans un groupe de N personnes il y a plus d'une chance sur deux pour qu'elles aient toutes des jours anniversaires différents.

Exercice : Dans une urne se trouvent 365 boules numérotées (autrement dit, « discernables »). On tire N fois une boule au hasard. Après chaque tirage on remet la boule. Quelle est la probabilité pour qu'on tire N boules différentes ?

Exercice Dans un établissement scolaire où les serrures sont nombreuses, on vous a prêté un trousseau de n clefs pour ouvrir une porte, et vous avez oublié de demander laquelle était la bonne. En plus, elles se ressemblent un peu toutes, impossible de deviner à vue laquelle est la plus plausible. Vous les essayez donc l'une après l'autre. Quelle est la probabilité pour que ce soit la $k$ième clé testée qui vous ouvre la porte ($1 \leq k \leq n$) ?

Définition (**Système complet d'événements**)

Soit (Ω, P) un espace probabilisé fini. On appelle système complet d'événements une famille $(A_i)_{1\leq i\leq m}$ d'événements deux à deux incompatibles dont la réunion est certaine, i.e. un famille de parties de Ω deux à deux disjointes et dont la réunion est Ω (ce qui signifie presque que $\{A_1, \ldots, A_m\}$ est une partition de Ω, à ceci près qu'il peut y avoir des A_i vides).

Par exemple, si A est un événement, (A, A^c) est un système complet d'événements.

Proposition (**Formule des probabilités totales**)

Soit $(A_i)_{1\leq i\leq m}$ un système complet d'événements tels que $\forall i \in \{1, \ldots m\} \quad P(A_i) > 0$. Alors, pour tout événement B,

$$P(B) \;=\; \sum_{k=1}^{m} P(A_k \cap B) \;=\; \sum_{k=1}^{m} P(A_k)\; P_{A_k}(B)$$

Proposition (**Formules de Bayes**)

Soit A et B deux événements tels que $P(A) > 0$ et $P(B) > 0$. Alors

$$P_B(A) = \frac{P(A)\; P_A(B)}{P(B)}$$

Si de plus $P(A^c) > 0$ alors

$$P_B(A) = \frac{P(A)\; P_A(B)}{P(A)\; P_A(B) \;+\; P(A^c)\; P_{A^c}(B)}$$

Et plus généralement, si $(A_i)_{1\leq i\leq m}$ est un système complet d'événements tels que $\forall i \in [\![1, m]\!] \quad P(A_i) \neq 0$, pour tout k on a

$$P_B(A_k) = \frac{P(A_k)\; P_{A_k}(B)}{\sum_{i=1}^{m} P(A_i) P_{A_i}(B)}$$

Démonstration de la formule des probabilités totales : Il suffit d'écrire

$$B = \bigcup_{k=1}^{m} (A_k \cap B)$$

et de remarquer que la réunion est disjointe. Il suffit alors d'utiliser l'additivité. On utilise cette formule pour décomposer le calcul de la probabilité d'un événement en plusieurs cas, soit par commodité, soit par nécessité. Pour la commodité, voir le deuxième exercice. Pour la nécessité, voir le premier exercice.

Démonstration des formules de Bayes : La première formule se déduit de

$$P(A \cap B) = P(A)P_A(B) = P(B)P_B(A)$$

les deux autres s'en déduisent par formule des probabilités totales.
Les applications typiques de la formule de Bayes se reconnaissent facilement (voir troisième exercice et suivants) et justifient l'appellation qui en est parfois donnée de « probabilité des causes ».

Exercice Pierre-Simon doit passer l'épreuve de Tipe des concours, et se demande s'il est judicieux d'utiliser son feutre vert pour les transparents. Son prof de math, pas avare il faut bien le dire de détails oiseux dans les nombreux conseils qu'il donne pour l'oral, a signalé que les personnes daltoniennes voyaient très mal cette couleur. Sur internet, on peut lire que 8% des hommes sont daltoniens, alors que seulement 0,5% des femmes le sont.
Par l'ami d'un cousin du voisin de son oncle, qui travaille dans l'administration du Tétraconcours, Pierre-Simon a appris que la moitié des jurys de l'épreuve de Tipe étaient composés de deux hommes, le quart d'un homme et d'une femme, le quart restant de deux femmes.
Quelle est la probabilité pour que les deux membres du jury que devra affronter Pierre-Simon soient daltoniens ?

Exercice On lance deux fois un dé équilibré. En conditionnant par le résultat du premier lancer, trouver la probabilité que la somme des deux résultats soit supérieure ou égale à 8.

Exercice On considère trois urnes U_1, U_2, U_3 contenant 12 boules chacune.
Dans U_1, il y a 4 boules rouges, 4 noires, 4 blanches.

Dans U_2, il y a 6 boules rouges, 2 noires, 4 blanches.
Dans U_3, il y a 2 boules rouges, 6 noires, 4 blanches.
Un individu jette un dé. Si le dé donne 1 ou 2, il tire une boule au hasard dans l'urne U_1. Si le dé donne 3 ou 4, il tire une boule au hasard dans l'urne U_2. Si le dé donne 5 ou 6, il tire une boule au hasard dans l'urne U_3. Un observateur (qui ne voit pas le résultat du lancer de dé, et ne peut pas distinguer les urnes, mais connaît les données du problème) constate que la boule tirée est rouge. S'il devait faire un pari, il aurait sûrement intérêt à miser sur le fait que le dé a donné 3 ou 4. Calculer la probabilité pour qu'effectivement le résultat du lancer de dé ait été un 3 ou un 4.

Exercice : Dans une population, 1 personne sur 10.000 est atteinte d'une maladie que l'on peut donc qualifier de relativement rare.
On dispose d'un test de détection de cette maladie. Des expérimentations ont permis de mesurer la sensibilité et la spécificité de ce test.
Sensibilité : la probabilité pour un individu malade d'être testé positif est 0,99.
Spécificité : la probabilité pour un individu non malade d'être testé négatif est 0,999.
On teste un individu pris au hasard dans la population. Le test est positif. Quelle est la probabilité pour qu'il soit malade ?

Exercice (supplément au précédent) Avec le même test et la même maladie, pour quelle proportion de malades dans la population arrive-t-on à la conclusion : « la probabilité pour qu'un individu testé positif soit malade est 0,9 » ? (on doit trouver le résultat suivant : une proportion d'environ 9 malades pour 1000 personnes).

Remarque : On considère trois urnes X, Y et Z. Dans l'urne X, il y a 99 boules noires, une boule blanche ; dans l'urne Y, il y a 999 boules blanches, une boule noire. Dans l'urne Z, il y a 9999 boules blanches, une boule noire. On tire au hasard une première boule dans l'urne Z. Si cette première boule est blanche, on tire une deuxième boule, dans l'urne Y. Sinon, on tire la deuxième boule

dans l'urne X. Quelle est la probabilité pour que la première boule soit noire sachant que la deuxième l'est ?

Exercice On lit dans un journal : « Alors qu'ils ne représentent que 13% de la population, les jeunes de 18 à 24 ans représentent 30% des tués sur les routes ». Rédiger une phrase équivalente à la précédente, sous la forme suivante : « un jeune de 18 à 24 ans a ...fois plus de chance de trouver la mort sur la route qu"une autre personne ».

Exercice Je viens d'emménager, je n'ai pas encore vu mes voisins, mais je sais qu'ils ont deux enfants.

1. Quelle est la probabilité que ces deux enfants soient des filles ?
2. J'ai vu un de ces deux enfants, c'est une fille ; même question.
3. On m'a dit que l'aînée de ces enfants est une fille ; même question.

On fait les hypothèses suivantes : à chaque naissance, la probabilité d'avoir une fille ou un garçon est la même. Les naissances sont indépendantes.

II.3 Indépendance

Définition On dit que deux événements A et B d'un espace probabilisé (Ω, P) sont indépendants lorsque

$$P(A \cap B) = P(A)\ P(B)$$

Remarquons que cela équivaut, dans le cas $P(B) > 0$, à

$$P_B(A) = P(A)$$

et donc aussi, dans le cas $P(A) > 0$, à

$$P_A(B) = P(B)$$

(ce sont ces deux dernières écritures qui constituent la définition raisonnable de l'indépendance : savoir si A est ou n'est pas réalisé ne nous donne aucune

information sur le fait que B le soit ou ne le soit pas. Mais la définition retenue a le double avantage d'être symétrique et de na pas faire d'hypothèse de non nullité des probabilités de A et B).
Lorsqu'on jette deux fois un dé, l'événement « le premier lancer donne un 2 » et l'événement « le deuxième lancer donne un 5 » seront en général, assez raisonnablement, considérés comme indépendants. Mais l'indépendance de deux événements n'est pas toujours fournie par une hypothèse "raisonnable" : elle peut être non intuitive et fournie par le calcul.
Il faut aussi éviter les confusions : l'indépendance n'est pas une notion « ensembliste » (si on change de probabilité sur l'univers Ω, la notion d'indépendance changera). Par exemple, on se convaincra facilement qu'il est très rare que deux événements disjoints soient indépendants...
Moins intuitive encore est l'indépendance de plus de deux événements :

Définition On dit que la famille d'événements $(A_i)_{1\leq i\leq m}$ est indépendante (ou que les A_i sont mutuellement indépendants) lorsque, pour toute partie finie I de $\{1,2,\ldots,m\}$,

$$P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i)$$

Pour que les A_i soient mutuellement indépendants, il est donc nécessaire qu'ils le soient deux à deux. Mais cela n'est pas suffisant. Considérons un exemple :
On lance un dé tétraédrique équilibré ; les faces sont marquées $1,2,3,4$. On considère les événements suivants :
A : le dé tombe sur la face 1 ou la face 2.
B : le dé tombe sur la face 2 ou la face 3.
C : le dé tombe sur la face 1 ou la face 3.
On vérifiera que ces événements sont deux à deux indépendants, mais ne le sont pas mutuellement.
Remarque 1 : Contrairement à ce qui se passe pour un dé cubique, il n'y a pas vraiment de face du dessus sur un dé tétraédrique. L'expression « le dé tombe sur la face i » doit donc être prise au pied de la lettre.
Remarque 2 : Le lemme de Novikov dit que pour une probabilité uniforme sur un ensemble de cardinal pq, où p et q sont deux nombres premiers distincts,

trois événements qui sont deux à deux indépendants le sont mutuellement. C'est le cas pour un lancer de dé équilibré à 6 faces. Si on n'aime pas les dés tétraédriques, on peut donc jouer deux fois à pile ou face, ou lancer deux fois un dé classique, on pourra alors trouver un contre-exemple analogue à celui donné ci-dessus.

Remarque : il ne faut donc pas comprendre « mutuellement » dans le sens « deux à deux ». On dit parfois que n variables aléatoires ($n > 2$) sont indépendantes « dans leur ensemble » pour signifier qu'elles le sont « mutuellement ». C'est peut-être un vocabulaire plus clair.

III Variables aléatoires

III.1 Définition ; loi d'une variable aléatoire

a. Un cadre et une difficulté : le jet de deux dés

On jette deux dés équilibrés. L'univers naturel est l'ensemble

$$\Omega = \{(i,j)\ ;\ 1 \leq i,j \leq 6\}$$

muni de la probabilité uniforme :

$$P\left(\{(i,j)\}\right) = \frac{1}{36}$$

On peut s'interroger sur cet univers ; pourquoi distinguer le résultat (5,6) et le résultat (6,5) ? lorsque les dés sont jetés (...), ces deux résultats sont indiscernables. Alors pourquoi ne pas prendre comme univers les paires $\{i,j\}$ (en acceptant $i = j$ pour un « double ») ?
Pourquoi pas, en effet. Mais la probabilité uniforme sur cet univers ne conviendrait pas à la modélisation du phénomène aléatoire considéré. La probabilité d'obtenir $\{6,6\}$ (un « double 6 ») n'est en effet pas la même que celle d'obtenir $\{5,6\}$ (« un 5 et un 6 »).
Penser que ces deux probabilités sont égales est un « biais cognitif » (plus simplement, une erreur très répandue, et pas seulement par ceux qui ne font pas de mathématiques) connu sous le nom de « biais d'équirépartition ».
On peut se convaincre assez aisément du fait qu'il est plus probable d'obtenir une paire $\{5,6\}$ qu'un double 6, par exemple en imaginant que les deux dés sont de couleurs différentes (s'il y a un dé rouge et un dé bleu, les événements « rouge donne 5, bleu donne 6 », « rouge donne 6, bleu donne 5 » et « rouge donne 6, bleu donne 6 » sont équiprobables).
Une autre manière de raisonner est de réfléchir au fait qu'il n'y a pas de différence entre le lancer simultané de deux dés et le lancer successif de deux dés.

On peut se tester en examinant le problème suivant :

> Deux joueurs A et B misent. Puis le joueur A lance une pièce. Si elle tombe sur Pile, il gagne. Sinon, il la relance. Si elle tombe sur Pile, il gagne. Sinon il perd, et c'est B qui gagne. Le gagnant emporte la mise.

Deux personnages célèbres, Diderot et D'Alembert, avaient réfléchi au problème suivant : combien doivent miser les deux joueurs pour avoir des chances égales ?
Pour que le jeu fût équitable, d'Alembert disait que A devait miser deux fois plus que B. Pour Diderot, A devait miser trois fois plus. Qui avait raison ?

b. Une variable aléatoire

Intéressons-nous à la somme S des deux dés : c'est une application

$$\begin{array}{rccc} S : & \Omega & \longrightarrow & E \\ & (i,j) & \longmapsto & i+j \end{array}$$

où on peut prendre $E = \mathbf{N}$, $E = \mathbf{R}$, ou, au plus juste, $E = [\![2,12]\!]$. Retenons ce dernier point de vue, inutile de rajouter des valeurs qui ne peuvent être atteintes par S. On calcule assez facilement :

k	$P(S=k)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

On peut alors définir sur $[\![2,12]\!]$ une probabilité P_S de la manière suivante :

$$\forall k \in [\![2,12]\!] \quad P_S(\{k\}) = P(S=k)$$

Ce qui permet de définir la probabilité de toute partie A de $[\![2,12]\!]$. Par exemple

$$P_S(\{6,7,8\}) = P(S \in \{6,7,8\}) = \frac{5}{36} + \frac{6}{36} + \frac{5}{36} = \frac{16}{36} = \frac{4}{9} \approx 0,444$$

Cette probabilité sur $[\![2,12]\!]$, qui est d'ailleurs tout ce qui nous intéresse si on veut parier sur la somme des deux dés, est appelée la loi de S. On remarquera que ce n'est pas du tout une loi uniforme.
Pourquoi appeler S une « variable aléatoire » ? parce qu'elle prend des valeurs aléatoires (à chaque issue de l'expérience aléatoire correspond une valeur). Ce qui nous importe, c'est la loi de S, c'est-à-dire la probabilité pour S de prendre telle ou telle valeur.

c. Définitions

Définition (variable aléatoire) Soit (Ω, P) un espace probabilisé fini. On appelle variable aléatoire sur (Ω, P) toute application X définie sur Ω à valeurs dans un ensemble E.

Définition (loi d'une variable aléatoire) Soit X une variable aléatoire définie sur l'espace probabilisé fini (Ω, P), à valeurs dans l'ensemble E. Pour toute partie A de E, on définit

$$P_X(A) = P\left(X^{-1}(A)\right)$$

où $X^{-1}(A) = \{\omega \in \Omega\,;\, X(\omega) \in A\}$ est aussi noté $\{X \in A\}$ ou $(X \in A)$. On a donc

$$P_X(A) = P(X \in A) = P(\{X \in A\})$$

P_X est la « loi » de X.

Proposition (propriétés de la loi d'une v.a.) Soit X une variable aléatoire définie sur l'espace probabilisé fini (Ω, P), à valeurs dans l'ensemble E. Soit P_X sa loi :

$$\begin{array}{rccc} P_X\ : & \mathscr{P}(E) & \longrightarrow & [0,1] \\ & A & \longmapsto & P(X \in A) \end{array}$$

On a $P(E) = 1$ et, pour toutes parties A et B de E,

$$(A \cap B = \emptyset) \Rightarrow (P_X(A \cup B) = P_X(A) + P_X(B))$$

Même si E n'est pas fini, on dira que P_X est une probabilité sur E.

Proposition (caractérisation de la loi d'une v.a.) Soit X une variable aléatoire définie sur l'espace probabilisé fini (Ω, P), à valeurs dans l'ensemble E. Remarquons que l'ensemble $X(\Omega) = \{X(\omega)\,;\, \omega \in \Omega\}$ est fini. Soit P_X la loi de X. Alors, si $A \in \mathscr{P}(E)$,

$$P_X(A) = \sum_{x \in X(\Omega) \cap A} P(X = x)$$

où l'événement $(X = x)$ est l'événement $(X \in \{x\})$. P_X est donc caractérisée par la donnée des $P(X = x)$ pour $x \in X(\Omega)$.

(on a d'ailleurs, pour l'exemple de la variable aléatoire S, fait le tableau de ces $P(S=k)$ pour connaître sa loi).

d. Variable en analyse et en probabilités

Lorsqu'on écrit $\sin x$, on appelle x « la variable » et sin est appelé « fonction » ou « application ».
Ici, quand on écrit $X(\omega)$, on appelle X « variable » (aléatoire certes, mais variable quand même...). Et ω ? on ne l'appelle pas...en fait, ω n'a pas vraiment d'importance : on remarquera qu'on parvient à l'éliminer des notations. Et en pratique, la détermination de l'univers Ω était souvent pénible, artificielle...et superflue. Ne reste donc que X et surtout le plus important : la **loi** de X. Quant au vocabulaire...quand x varie, son sinus varie aussi. Les deux sont donc fondés à réclamer le titre de variable.

e. Exemple-exercice : une urne de Polya

Ici, plutôt que de vider une urne, on va la remplir...
Soit une urne contenant une boule blanche et une boule noire. On dispose par ailleurs d'une réserve suffisante de boules blanches et noires, et on répète n fois l'action suivante :
On tire une boule dans l'urne. Si cette boule est blanche (respectivement noire), on la remet dans l'urne, et on rajoute une boule blanche (respectivement noire) dans l'urne.
On désigne par X_n la variable aléatoire « nombre de boules blanches dans l'urne » après ces n tirages. Quelle est la loi de X_n ?

Indication : examiner les petites valeurs de n peut donner des idées

III.2 Espérance d'une variable aléatoire réelle

a. Définition

Définition Soit X une variable définie sur un espace probabilisé fini (Ω, P), à valeurs réelles. On définit l'espérance de X :

$$\mathrm{E}(X) = \sum_{\omega\in\Omega} P(\{\omega\})\ X(\omega)$$

Ainsi, l'espérance de X est la moyenne des valeurs prises par X pondérée par les probabilités d'obtention de ces valeurs. Si la loi de X est uniforme, cette espérance est la moyenne arithmétique des valeurs prises par X. Dans tous les cas, c'est un barycentre. Et c'est un nombre réel.

Réécriture (transfert) Soit X une variable définie sur un espace probabilisé fini (Ω, P), à valeurs réelles. Soit $A = \{a_1, \dots, a_n\} \subset \mathbf{R}$ un ensemble fini contenant $X(\Omega)$. Alors :

$$\mathrm{E}(X) = \sum_{k=1}^{n} P(X = a_k)\ a_k$$

C'est bien sûr cette formule qu'on utilise pour le calcul pratique : on voit souvent qu'on s'intéresse si peu à l'univers Ω qu'on ne prend pas la peine de le définir. La première définition semble alors peu utilisable ! On a donc « transféré » le problème, de (Ω, P) vers (A, P_X), où A est une partie de $\mathbf{R}$ qui contient $X(\Omega)$.

Démonstration Pour tout k,

$$P(X = a_k) = P\left(X^{-1}(\{a_k\}\right) = \sum_{\omega\in X^{-1}(\{a_k\}} P(\{\omega\})$$

Et donc

$$\sum_{k=1}^{n} a_k P(X = a_k) = \sum_{k=1}^{n} \left(\sum_{\omega\in X^{-1}(\{a_k\}} X(\omega) P(\{\omega\}) \right)$$

(car, si $\omega \in X^{-1}(\{a_k\})$, $X(\omega) = a_k$). Et, comme les a_k sont supposés distincts, Ω est la réunion disjointe des $X^{-1}(\{a_k\})$ (certains de ces ensembles pouvant être vides sans que cela soit gênant). Le résultat s'ensuit.

Exemple : Reprenons la variable S somme de deux dés vue plus haut. On a

$$\mathrm{E}(S) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + \ldots + 11 \times \frac{2}{36} + 12 \times \frac{1}{36} = \frac{252}{36} = 7$$

(calculer en regroupant les termes en k et $14-k$ dans la somme).

Exercice : Dans une fête foraine, pour une mise initiale de 3 euros, le joueur est invité à lancer deux dés (que l'on supposera équilibrés). Si le résultat est un double, le joueur récupère sa mise et empoche en plus le montant (en euros) égal à la somme des points obtenus. Si un dé et un seulement fait apparaître un 6, le joueur récupère sa mise et gagne en le montant en euros indiqué sur l'autre dé. Sinon, la partie est perdue (et avec elle, les trois euros...). Calculer l'espérance de G, gain (en euros) du joueur (bien entendu, une perte est comptée comme un gain négatif).

Solution : Soit $i \in [\![1,6]\!]$. L'événement « apparition d'un double i » a pour probabilité $1/36$. Et à cet événement correspond une valeur de G égale à i.
Soit $i \in [\![1,5]\!]$; l'événement « apparition d'un 6 et d'un i » a pour probabilité $2/36$ et à cet événement correspond un gain $G = i$. Dans tous les autres cas, $G = -3$. L'espérance cherchée est donc

$$\mathrm{E}(G) = \frac{1}{36}\left(\sum_{i=1}^{6} i \; + \; 2\sum_{i=1}^{5} i \; + \; 20 \times (-3)\right) = \frac{-1}{4}$$

Remarque 1 : Pour $i = 1, 2, \ldots, 5$, on n'a pas directement calculé la probabilité de l'événement $G = i$. On a partitionné cet événement en deux sous-événements (apparition d'un double i, apparition d'un i et d'un 6).
Remarque 2 : L'espérance de gain (-25 centimes) justifie dans ce cadre son nom (donné par Huyghens, en latin : expectatio). On peut en effet justifier mathématiquement l'intuition suivante : si on joue un « grand nombre » (n) de fois, on

aura « de bonnes chances » de perdre une somme « voisine » de $n \times 25$ centimes. Tout cela demande bien sûr à être précisé. Mais les lois des grands nombres et l'espérance de gain positive pour lui permettent au forain de gagner (modestement) sa vie. Alors qu'il suffit d'une probabilité non négligeable de gain intéressant pour que le client accepte de jouer.

Définition On dit qu'une variable aléatoire est centrée lorsqu'elle a une espérance nulle.

b. Propriétés de l'espérance

Proposition *Les propriétés qui suivent rappellent quelque chose : l'espérance ne serait-elle pas une intégrale ? si, mais le point de vue de Riemann n'est pas idéal pour développer cette idée.*

- Si X et Y sont deux variables aléatoires réelles, si α et β sont deux réels,
$$\mathrm{E}(\alpha X + \beta Y) = \alpha \mathrm{E}(X) \ + \ \beta \mathrm{E}(Y)$$
(linéarité)
- Si X est une variable aléatoire positive (i.e. X est à valeurs dans $\mathbf{R}^+$), alors
$$\mathrm{E}(X) \geq 0$$
(positivité)
- Si X et Y sont deux variable aléatoire réelles telles que $X \leq Y$ (i.e. $\forall \omega \in \Omega \quad X(\omega) \leq Y(\omega)$), alors
$$\mathrm{E}(X) \leq \mathrm{E}(Y)$$
(croissance)
- Si X est une variable aléatoire réelle, alors
$$|\mathrm{E}(X)| \ \leq \mathrm{E}(|X|)$$
- Si X est constante : $\forall \omega \in \Omega \quad X(\omega) = a$, alors
$$\mathrm{E}(X) = a$$

III.3 Variance et écart-type

a. Définitions

On a défini l'espérance d'une variable aléatoire comme une valeur moyenne. On peut essayer de quantifier la « dispersion » ou « variation » de X autour de cette valeur moyenne. Une idée serait de considérer

$$E(|X - \mathrm{E}(X)|)$$

mais ce n'est pas très maniable et il y a mieux. Si on se souvient que la norme euclidienne est plus agréable que $\|.\|_1$ (car elle est associée à un produit scalaire), on comprendra que l'on préfère :

Définition Soit X une variable aléatoire réelle sur un espace probabilisé fini (Ω, P). On définit sa variance :

$$\mathrm{V}(X) = \mathrm{E}\left((X - \mathrm{E}(X))^2\right)$$

($X - \mathrm{E}(X)$ est la variable aléatoire $\omega \to X(\omega) - E(X)$). On remarque que $V(X) \geq 0$, ce qui permet de définir l'écart-type de X :

$$\sigma(X) = \sqrt{V(X)}$$

On a alors une propriété simple :

Proposition Une variable aléatoire sur un espace probabilisé fini a une variance nulle si et seulement si elle est constante.

Remarque : Si a est un nombre réel, on note aussi a la variable aléatoire constante égale à a (notation déjà utilisée pour $a = E(X)$, qui est un nombre mais est aussi considéré comme une variable aléatoire constante).

Remarque : L'application $(X, Y) \mapsto \mathrm{E}(XY)$ est un produit scalaire sur l'espace des variables aléatoires réelles définies sur l'espace probabilisé fini (Ω, P) (démonstration sans difficulté). Si on note $\|.\|$ la norme euclidienne associée, et d la distance, on a

$$\sigma(X) = \|X - \mathrm{E}(X)\| = d(X, E(X))$$

b. Deux formules

$$\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$$

$$\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$$

(a et b sont des constantes réelles).

Démonstration
Pour la première formule, X est une variable aléatoire réelle sur un espace probabilisé fini. Pour plus de clarté, il est judicieux de renommer l'espérance : notons $\lambda = \mathrm{E}(X)$. Alors

$$\begin{aligned}\mathrm{V}(X) &= \mathrm{E}\left((X-\lambda)^2\right)\\ &= \mathrm{E}\left(X^2 - 2\lambda X + \lambda^2\right)\\ &= \mathrm{E}(X^2) - 2\lambda\mathrm{E}(X) + \lambda^2\\ &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2\end{aligned}$$

(On a utilisé la linéarité de l'espérance, et le fait que l'espérance d'une variable aléatoire constante est égale à cette constante) (on ne sera pas sans remarquer le curieux assortiment de mots figurant dans l'expression « variable aléatoire constante », tant une constante est peu variable et pas si aléatoire que ça).

Pour la deuxième formule, on peut par exemple utiliser la première :

$$\begin{aligned}\mathrm{V}(aX+b) &= \mathrm{E}\left((aX+b)^2\right) - (\mathrm{E}(aX+b))^2\\ &= \mathrm{E}\left(a^2X^2 + 2abX + b^2\right) - (a\mathrm{E}(X)+b)^2\\ &= a^2\mathrm{E}(X^2) + 2ab\mathrm{E}(X) + b^2 - a^2\left(\mathrm{E}(X)\right)^2 - 2ab\mathrm{E}(X) - b^2\\ &= a^2V(X)\end{aligned}$$

(on utilise les mêmes propriétés que dans la démonstration précédente)

Rappel : Contrairement aux usages courants en analyse, on note ici de la même manière un nombre réel et la « variable aléatoire » (fonction) constante qui prend pour unique valeur ce nombre réel.

Corollaire

$$(\mathrm{E}(X))^2 \leq \mathrm{E}(X^2)$$

Démonstration, et considérations euclidiennes L'inégalité résulte directement de la première formule et de la positivité de la variance, et c'est la démonstration la plus courte, mais il est intéressant de la voir autrement (si vous avez reconnu une ressemblance avec Cauchy-Schwarz, bravo !) :
Rappelons que

$$\langle X, Y\rangle = \mathrm{E}(XY)$$

définit (c'est très simple à vérifier) un produit scalaire sur l'espace des variables aléatoires réelles définies sur l'espace probabilisé fini (Ω, P). On note $\|.\|$ la norme euclidienne associée : $\|X\| = \sqrt{\mathrm{E}(X^2)}$. L'inégalité de Cauchy-Schwarz s'écrit alors :

$$|\mathrm{E}(XY)| \leq \sqrt{\mathrm{E}(X^2)}\sqrt{\mathrm{E}(Y^2)}$$

et, prenant Y constante et égale à 1 :

$$|\mathrm{E}(X)| \leq \sqrt{\mathrm{E}(X^2)}$$

qui élevée au carré donne bien l'inégalité voulue.
Continuons à exploiter ce point de vue euclidien, et cherchons la projection orthogonale de X sur le sous-espace des (variables)(fonctions) constantes ; cette projection a est caractérisée par

$$\langle X - a, 1\rangle = 0$$

(car la (variable) constante 1 engendre le sous-espace vectoriel des constantes). Ce qui équivaut à $\mathrm{E}(X - a) = 0$, ou encore $a = \mathrm{E}(X)$. L'espérance de la variable aléatoire X peut donc être vue comme sa projection orthogonale sur le sous-espace des constantes. On peut alors (faire un dessin) appliquer le théorème de Pythagore :

$$\mathrm{E}(X^2) = \mathrm{E}\left((\mathrm{E}(X))^2\right) + \mathrm{E}\left((X - \mathrm{E}(X))^2\right) = (\mathrm{E}(X))^2 + V(X)$$

et on retrouve la première formule sur la variance.
De plus $\sigma(X)$ est la distance euclidienne de X au sous-espace des constantes. Donc $\sigma(aX + b) = \sigma(aX)$ ce qui permet de retrouver rapidement la deuxième formule ($\mathrm{V}(aX) = a^2\mathrm{V}(X)$ est assez simple).

c. Variable centrée réduite

On dit que la variable aléatoire X est « centrée réduite » lorsque

$$\mathrm{E}(X) = 0 \quad \text{et} \quad \mathrm{V}(X) = 1$$

Soit X une variable aléatoire de variance non nulle. Notons, comme d'habitude, $m = \mathrm{E}(X)$, $\sigma = \sqrt{\mathrm{V}(X)}$ Alors la variable aléatoire

$$Y = \frac{X - m}{\sigma}$$

est centrée réduite. Autrement dit, on peut écrire

$$X = m + \sigma Y$$

où Y est centrée réduite.

Démonstration Utilisation simple de formules déjà vues.

IV Lois usuelles

Quelques lois classiques sont à connaître, avec leurs espérances et leurs variances.

IV.1 Loi uniforme

Soit X une variable aléatoire, définie sur un espace probabilisé fini (Ω, P), et soit $E = X(\Omega) = \{x_1, \dots, x_n\}$. On dira que X suit une loi uniforme lorsque P_X est la probabilité uniforme sur $\{x_1, \dots, x_n\}$:

$$\forall k \in [\![1, n]\!] \qquad P(X = x_k) = \frac{1}{n}$$

L'espérance et la variance de X sont alors :

$$\mathrm{E}(X) = \frac{x_1 + \cdots + x_n}{n} = \overline{x}$$

$$\mathrm{V}(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

IV.2 Loi de Bernoulli

a. Définition

On joue à Pile ou Face, avec une pièce pas forcément équilibrée. La variable aléatoire qui prend la valeur 1 si le résultat du lancer est Pile et 0 si c'est face suit une loi de Bernoulli :

Définition On appelle variable de Bernoulli toute variable aléatoire X à valeurs dans $\{0,1\}$; sa loi est donc donnée par

$$P(X=1)=p \qquad P(X=0)=1-p$$

où $p \in [0,1]$. On dit que X suit une loi de Bernoulli de paramètre p, ou que X suit une loi $\mathscr{B}(p)$.

Espérance, variance Si X suit une loi $\mathscr{B}(p)$, alors

$$\mathrm{E}(X)=p \qquad \mathrm{V}(X)=p(1-p)=pq$$

où l'on note $q=1-p$.

Notons que la loi $\mathscr{B}(1/2)$ est une loi uniforme.
La loi de Bernoulli modélise les expériences aléatoires à deux issues, dont l'une est souvent appelée « succès » et l'autre « échec ».

b. Variable de Bernoulli et fonction indicatrice d'un événement

Soit (Ω, P) un espace probabilisé fini. Si A est un événement quelconque (autrement dit : une partie de Ω), on considère sa fonction indicatrice, qu'on appelle plus brièvement son indicatrice : $\mathbf{1}_A$ définie par

$$\forall \omega \in A \quad \mathbf{1}_A(\omega)=1 \qquad \text{et} \qquad \forall \omega \in \Omega \setminus A \quad \mathbf{1}_A(\omega)=0$$

$\mathbf{1}_A$ est une variable aléatoire, et suit une loi $\mathscr{B}(p)$ avec $p=P(A)$. Réciproquement, toute variable de Bernoulli X est l'indicatrice de l'événement $\{X=1\}$.

IV.3 Loi binomiale

a. Définition

On joue encore à Pile ou Face, avec une pièce pas forcément équilibrée. On suppose que la probabilité de tomber sur Pile est p, et on note $q=1-p$. On lance n fois la pièce ($n \in \mathbf{N}_*$), et on désigne par X le nombre fois où on obtient Pile. Détaillons :

On peut considérer l'univers $\Omega = \{0,1\}^n$ (ensemble des suites finies de n entiers égaux à 0 ou 1), on représentera par exemple Pile par 1 et Face par 0. Une issue (ou réalisation) de l'expérience est donc une suite de n entiers égaux à 0 ou 1. On peut représenter une telle suite par un n-uplet.
L'événement $(X = k)$ est l'ensemble des n-uplets dans lesquelles « 1 » figure k fois et « 0 » $n-k$ fois. Il y a $\binom{n}{k}$ tels événements élémentaires (nombre de n-uplets comportant k « 1 » et $n-k$ « 0 », chacun de ces n-uplets se caractérisant par les places occupées par les k « 1 »).
La probabilité de chacune de ces réalisations est $p^k(1-p)^{n-k}$ (formule des probabilités composées) car le résultat de chaque lancer est implicitement (et raisonnablement !) supposé indépendant des autres. Finalement,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Définition On dit qu'une variable aléatoire suit une loi binomiale de paramètres $n \in \mathbf{N}_*$ et $p \in [0,1]$, ou qu'elle suit une loi $\mathscr{B}(n,p)$, lorsqu'elle est à valeurs dans $[\![0,n]\!]$ et vérifie

$$\forall k \in [\![0,n]\!] \quad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

La loi $\mathscr{B}(n,p)$ est donc la loi du nombre de succès dans l'expérience consistant à répéter n fois des expériences de Bernoulli indépendantes de paramètre p.
C'est aussi la loi de la somme de n variables indépendantes de loi $\mathscr{B}(p)$ (voir plus loin). La loi de Bernoulli est une loi binomiale particulière ($n = 1$).

b. Espérance, variance

« La » bonne idée est de déduire ces résultats des résultats sur l'espérance et la variance d'une somme de variables aléatoires indépendantes, mais faire les calculs directement est un intéressant exercice de sommation. On considère

donc une variable aléatoire X qui suit une loi $\mathscr{B}(n,p)$:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
&= n\sum_{k=1}^{n}\frac{k}{n}\binom{n}{k}p^k(1-p)^{n-k} \\
&= n\sum_{k=1}^{n}\binom{n-1}{k-1}p^k(1-p)^{n-k} \\
&= np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j(1-p)^{n-1-j} \\
&= np(p+1-p)^{n-1} \\
&= np
\end{aligned}$$

(avec une réindexation $j = k-1$). De la même manière :

$$\begin{aligned}
\mathrm{E}(X^2) &= \sum_{k=0}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k} \\
&= n\sum_{k=1}^{n} k\frac{k}{n}\binom{n}{k}p^k(1-p)^{n-k} \\
&= n\sum_{k=1}^{n} k\binom{n-1}{k-1}p^k(1-p)^{n-k} \\
&= n\sum_{k=1}^{n}(k-1+1)\binom{n-1}{k-1}p^k(1-p)^{n-k} \\
&= n(n-1)\sum_{k=1}^{n}\frac{k-1}{n-1}\binom{n-1}{k-1}p^k(1-p)^{n-k} \;+\; np \\
&= n(n-1)\sum_{j=0}^{n-2}\binom{n-2}{j}p^{j+2}(1-p)^{n-2-j} \;+\; np \\
&= n(n-1)p^2 + np
\end{aligned}$$

et donc

$$\mathrm{V}(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq$$

Il y a une méthode plus astucieuse pour réaliser ces sommations : calculer de deux manières différentes la dérivée et la dérivée seconde en 0 de

$$t \mapsto (1-x+xe^t)^n$$

(d'une part en dérivant sous cette forme, d'autre part en développant par la formule du binôme puis en dérivant). C'est beaucoup plus rapide, mais très astucieux, la bonne méthode reste évidemment d'utiliser l'interprétation d'une variable $\mathscr{B}(n,p)$ comme somme de n variables $\mathscr{B}(p)$ indépendantes (voir plus loin). Il faut retenir :

Si X suit une loi $\mathscr{B}(n,p)$, alors

$$\mathrm{E}(X) = np \qquad \mathrm{V}(X) = np(1-p) = npq$$

où l'on note $q = 1-p$.

V Inégalités classiques

Deux inégalités très simples : l'inégalité de Markov, qui est une inégalité intégrale (mais notre cadre d'intégrale de Riemann ne rend pas très naturelle l'interprétation de l'espérance comme intégrale), et l'inégalité de Bienaymé-Tchebychev qui en est une conséquence.
Simples, et pourtant déjà bien utiles. L'inégalité de Bienaymé-Tchebychev permet par exemple de montrer la loi « faible » des grands nombres.

V.1 Inégalité de Markov

Proposition (**Inégalité de Markov**)

Soit X une variable aléatoire réelle définie sur un espace probabilisé fini (Ω, P). Alors, pour tout réel strictement positif a,

$$P(|X| \geq a) \leq \frac{\mathrm{E}(|X|)}{a}$$

Il est assez banal de dire que cette inégalité majore la probabilité pour une variable aléatoire de prendre de grandes valeurs (« grandes » se comprenant en valeur absolue). Elle peut aussi être écrite :

Proposition (**Inégalité de Markov**)

Soit X une variable aléatoire réelle définie sur un espace probabilisé fini (Ω, P). Alors, pour tous réels strictement positifs a et k,

$$P(|X| \geq a) \leq \frac{\mathrm{E}(|X|^k)}{a^k}$$

Démonstration Il s'agit de montrer que

$$aP(|X| \geq a) \leq \mathrm{E}(|X|)$$

ce qui est « évident » si on voit l'espérance comme une intégrale, mais passons.

Une première démonstration

Une idée est d'écrire

$$\mathrm{E}(|X|) = \sum_{k=1}^{n} P\left(|X| = \alpha_k\right)\ \alpha_k$$

où $|X|(\Omega) = \{\alpha_1, \ldots, \alpha_n\}$; on considère alors $I = \{k \in [\![1, n]\!]\ ;\ \alpha_k \geq a\}$. On a

$$\begin{aligned}\mathrm{E}(|X|) &\geq \sum_{k\in I} P\left(|X| = \alpha_k\right)\ \alpha_k \\ &\geq a \sum_{k\in I} P\left(|X| = \alpha_k\right) \\ &= aP(|X| \geq a)\end{aligned}$$

(l'événement $\{|X| \geq a\}$ étant la réunion disjointe des événements $\{|X| = \alpha_k\}$, $k \in I$).

Une deuxième démonstration

Cette démonstration peut paraître moins élémentaire, voire astucieuse, mais elle ne l'est pas lorsqu'on voit l'espérance comme une intégrale (décidément…), elle vaut la peine d'être comprise.

Notons $A = \{|X| \geq a\} = \{\omega \in \Omega\ ;\ |X(\omega)| \geq a\}$. On a

$$a\mathbf{1}_A \leq |X|\ \mathbf{1}_A \tag{1}$$

Rappelons que $\mathbf{1}_A$ est l'indicatrice de A, c'est une variable aléatoire qui vaut 1 sur A et 0 ailleurs. L'inégalité (1), qui dit que pour tout $\omega \in \Omega$ on a $a\mathbf{1}_A(\omega) \leq |X(\omega)|\mathbf{1}_A(\omega)$, se vérifie alors en distinguant deux cas : $\omega \in A$ et $\omega \notin A$. Cette inégalité (1) est une inégalité entre variables aléatoires ; on la vérifie en distinguant deux cas : $\omega \in A$ et $\omega \notin A$. Par croissance :

$$\mathrm{E}\left(a\mathbf{1}_A\right) \leq \mathrm{E}\left(|X|\ \mathbf{1}_A\right) \tag{2}$$

Mais $\mathrm{E}\left(a\mathbf{1}_A\right) = a\mathrm{E}\left(\mathbf{1}_A\right) = aP(A) = aP(|X| \geq a)$. Et $|X|\mathbf{1}_A \leq |X|\mathbf{1}_\Omega = |X|$, d'où, par croissance encore,

$$aP(|X| \geq a) \leq \mathrm{E}\left(|X|\ \mathbf{1}_A\right) \leq \mathrm{E}(|X|)$$

ce qui conclut. En remplaçant a par a^k et $|X|$ par $|X|^k$ on obtient la généralisation annoncée, la fonction $x \mapsto x^k$ étant croissante sur $\mathbf{R}^+$.

V.2 Inégalité de Bienaymé-Tchebychev

Une des plus importantes conséquences de l'inégalité de Markov est l'inégalité de Bienaymé-Tchebychev :

Proposition (Inégalité de Bienaymé-Tchebychev)

Soit X une variable aléatoire réelle définie sur un espace probabilisé fini (Ω, P). Alors, pour tout réel strictement positif a,

$$P(|X - \mathrm{E}(X)| \geq a) \leq \frac{\mathrm{V}(X)}{a^2}$$

Démonstration On applique l'inégalité de Markov, avec $p = 2$, à la variable aléatoire $X - \mathrm{E}(X)$. Cette inégalité majore la probabilité de s'écarter de la moyenne. Sa conséquence la plus célèbre est la loi faible des grands nombres...et c'est en l'étudiant que Bernstein a trouvé sa démonstration du premier théorème de Weierstrass.
Remarque : L'orthographe de Tchebychev est très fluctuante dans la littérature mathématique.

Exercice :

1. On lance 1000 fois une pièce équilibrée ; minorer la probabilité pour que le nombre de "Pile" soit compris entre 450 et 550.
2. On lance n fois une pièce équilibrée. On note Z le nombre de "Pile". Trouver n pour que
$$P\left(\left|\frac{Z}{n} - \frac{1}{2}\right| \leq 0,05\right) \geq 0,99$$

On demande un n qui convient, pas le plus petit.

VI Image d'une variable aléatoire par une fonction

VI.1 Définition-Notation

Soit X une variable aléatoire sur un espace probabilisé fini (Ω, P), à valeurs dans un ensemble E. Soit F un autre ensemble (éventuellement le même d'ailleurs, ça n'est pas gênant), et ϕ une application définie sur une partie de E contenant $X(\Omega)$, à valeurs dans F.

$$\Omega \xrightarrow{X} E \xrightarrow{\phi} F$$

On peut donc définir sur Ω la variable aléatoire

$$Y = \phi \circ X \ : \ \omega \mapsto \phi(X(\omega))$$

Y est la composée de X par ϕ. Mais on dira plutôt que Y est l'image de X par ϕ. Et on notera $Y = \phi(X)$ plutôt que $Y = \phi \circ X$.

Sans en parler, on a déjà considéré ce genre d'objet : si X est une variable aléatoire réelle, X^2 est l'image de X par la fonction $x \mapsto x^2$, $|X|$ est l'image de X par la fonction valeur absolue.

VI.2 Loi

On reprend la définition précédente. La loi P_Y de Y est définie sur $\mathscr{P}(F)$ par

$$\begin{array}{rccc} P_Y \ : & \mathscr{P}(F) & \longrightarrow & [0,1] \\ & B & \longmapsto & P(Y \in B) \end{array}$$

mais, bien évidemment, on veut comme d'habitude se débarasser de Ω...et, tant qu'à faire, de P, d'autant que ce qu'on connaît en pratique, c'est la loi P_X de X, et non P.

On remarque donc que, pour toute partie B de F,

$$\begin{aligned} P_Y(B) &= P(Y \in B) \\ &= P\left(\phi \circ X \in B\right) \\ &= P\left(X \in \phi^{-1}(B)\right) \\ &= P_X\left(\phi^{-1}(B)\right) \end{aligned}$$

On a donc

$$\forall B \in \mathscr{P}(F) \qquad P_Y(B) = P_X\left(\phi^{-1}(B)\right)$$

Cela dit, l'ensemble des valeurs prises par Y est fini ; si y est une valeur prise par Y, on a

$$P(Y=y) = P\left(X \in \phi^{-1}(\{y\})\right) = \sum_{x \ / \ \phi(x)=y} P(X=x) = \sum_{x \in \phi^{-1}(\{y\})} P_X(\{x\}) = \sum_{x \in \phi^{-1}(\{y\})} P(X=x)$$

VI.3 Espérance (formule de transfert)

Soit X une variable aléatoire définie sur un espace probabilisé fini (Ω, P), à valeurs dans un ensemble E. Soit f une fonction définie au moins sur $X(\Omega)$, à valeurs réelles. Alors

$$\mathrm{E}\left(f(X)\right) = \sum_{x \in X(\Omega)} P(X=x)\ f(x)$$

Cette formule transfère le calcul de l'espérance de (Ω, P) vers (E, P_X), et donc est fondamentale car on ne formalise en général pas (Ω, P) (pas unique, trop compliqué, pas utile...).

Démonstration :
Notant $Y = f(X)$,

$$E(Y) = \sum_{y \in Y(\Omega)} P(Y=y)\,y$$

Or $Y(\Omega) = \{f(x)\ ;\ x \in X(\Omega)\}$. Pour tout $y \in Y(\Omega)$, notons $A_y = \{x \in X(\Omega)\ ;\ f(x) = y\}$. Alors

$$P(Y=y) = P(f(X)=y) = \sum_{x \in A_y} P(X=x)$$

(l'événement $f(X) = y$ est réunion disjointe des événements $X = x$ pour $x \in A_y$). et donc

$$\begin{aligned} E(Y) &= \sum_{y \in Y(\Omega)} \left(\sum_{x \in A_y} P(X = x) f(x) \right) \\ &= \sum_{x \in X(\Omega)} P(X = x)\ f(x) \end{aligned}$$

car $X(\Omega)$ est réunion disjointe des A_y pour $y \in Y(\Omega)$.

Exercice : On suppose que X suit une loi binomiale de paramètres n et p, $n \in \mathbf{N}_*$, $p \in [0,1]$(on note : $X \sim \mathscr{B}(n,p)$). Calculer l'espérance de la variable aléatoire

$$Y = \frac{1}{1+X}$$

VII Couple de variables aléatoires

VII.1 Loi conjointe, lois marginales

a. Exemple

On lance deux fois un dé équilibré. On considère les variables aléatoires X, résultat du premier lancer, à valeurs dans $E = \{1,2,3,4,5,6\}$, et S, somme des résultats des deux lancers, à valeurs dans $F = \{2,3,\dots,12\}$. Le couple (X,S) est ainsi une variable aléatoire, à valeurs dans $E \times F$. Si $(x,s) \in E \times F$, on notera $(X = x, S = s)$ l'événement $(X,S) = (x,s)$, qui n'est autre que l'événement $(X = x) \cap (S = s)$ que l'on écrit aussi ($X = x$ et $S = s$).
Dans le tableau ci-dessous, à l'intersection de la ligne x et de la colonne s on trouve $P(X = x, S = s)$:

	2	3	4	5	6	7	8	9	10	11	12
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36

On a ainsi décrit la loi du couple (X,S), appelée « loi conjointe ». Si A est une partie de $[\![1,6]\!]$ et B une partie de $[\![2,12]\!]$, on pourra calculer

$$\begin{aligned} P(X \in A, S \in B) &= P(X \in A \text{ et } S \in B) \\ &= P\left(\bigcup_{(x,s)\in A\times B} P(X = x \text{ et } S = s) \right) \\ &= \sum_{(x,s)\in A\times B} P(X = x \text{ et } S = s) \end{aligned}$$

Par exemple, en ajoutant des 0 et des 1/36, le tableau ci-dessus permet de calculer

$$P(2 \leq X \leq 4 \text{ et } 6 \leq S \leq 8) = 1/4$$

(en fait, en l'occurrence il n'y a pas tellement de 0 à ajouter...).
Faisons la somme des nombres inscrits dans chaque colonne ; on obtient (en première ligne le rappel de la valeur de s, en deuxième ligne la somme des probabilités figurant dans la colonne correspondante) :

2	3	4	5	6	7	8	9	10	11	12
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

...et on retrouve la loi de S (déjà vue plus haut). Comme cette somme se fait "naturellement" dans la marge du bas du tableau (ou dans celle du haut), on l'appellera loi marginale. De même, dans la marge de droite (ou de gauche) du tableau, on reconstituerait la loi de X en faisant la somme de chaque ligne (moins intéressante : on trouve $1/6$ sur chaque ligne, X suit une loi uniforme).

b. Définitions

Le vocabulaire « conjointe » et « marginale » désigne l'aller et retour entre les lois de X et de Y d'une part, la loi de (X,Y) d'autre part.

Définition Soit X, Y deux variables aléatoires définies sur un espace probabilisé fini (Ω, P), à valeurs dans E et F respectivement. La « loi conjointe » de X et Y est la loi du couple (X,Y), qui est une variable aléatoire à valeurs dans $E \times F$.

Définition Avec les notations précédentes, les lois marginales du couple (X,Y) sont les lois de X et de Y.

On a vu sur un exemple que les lois marginales se déduisaient facilement de la loi conjointe : si $X(\Omega) = \{x_1, \dots, x_p\}$ et $Y(\Omega) = \{y_1, \dots, y_q\}$, on remarque que $(Y = y_j)_{1 \leq j \leq q}$ est un système complet d'événements, ce qui permet d'écrire

$$\forall i \in [\![1, p]\!] \qquad P(X = x_i) = \sum_{j=1}^{q} P(X = x_i \text{ et } Y = y_j)$$

et, de même,

$$\forall j \in [\![1, q]\!] \qquad P(Y = y_j) = \sum_{i=1}^{p} P(X = x_i \text{ et } Y = y_j)$$

En revanche, il est illusoire de penser reconstituer le tableau à partir de ses "marges" : autrement dit,

Remarque importante : Les lois marginales ne déterminent pas les lois conjointes

Par exemple,

X \ Y	α	β
a	1/4	1/4
b	1/4	1/4

et

X \ Y	α	β
a	1/6	1/3
b	1/3	1/6

donnent les mêmes lois marginales (loi uniforme sur $\{a,b\}$ pour X, loi uniforme sur $\{\alpha,\beta\}$ pour Y), mais pas la même loi conjointe : dans le premier cas, la loi de (X,Y) est la probabilité uniforme sur $\{a,b\}\times\{\alpha,\beta\}$, dans le second cas la loi de (X,Y) n'est pas uniforme.

VII.2 Loi conditionnelle

Définition Soit X, Y deux variables aléatoires définies sur un espace probabilisé fini (Ω,P), à valeurs dans E et F respectivement. Soit $x\in E$ tel que $P(X=x)>0$. On définit la loi conditionnelle de Y sachant $(X=x)$ par

$$\forall B\in\mathscr{P}(F)\qquad P_{Y|X=x}(B)=P_{X=x}(Y\in B)=\frac{P(Y\in B \text{ et } X=x)}{P(X=x)}$$

Cette « loi conditionnelle », notée $P_{Y|X=x}$, est assez clairement une probabilité (voir probabilités conditionnelles).
La loi conditionnelle de Y sachant $(X=x)$ se caractérise par les probabilités des événements élémentaires :

$$P_{Y|X=x}(\{y\})=P_{X=x}(Y=y)=\frac{P(Y=y \text{ et } X=x)}{P(X=x)}$$

et, ensuite, pour toute partie B de F,

$$P_{Y|X=x}(B)=\sum_{y\in B\cap Y(\Omega)}P_{X=x}(Y=y)=\sum_{y\in B\cap Y(\Omega)}\frac{P(Y=y \text{ et } X=x)}{P(X=x)}$$

VII.3 Couple de variables aléatoires indépendantes

a. Définition

Proposition Soit X, Y deux variables aléatoires définies sur un espace probabilisé fini (Ω, P), à valeurs dans E et F respectivement. On note P_X, P_Y leurs lois. Alors il y a équivalence entre les trois propriétés suivantes :

(i) Pour tout $x \in E$ tel que $P(X = x) > 0$, la loi conditionnelle de Y sachant $X = x$ est celle de Y (i.e. $P_{Y|X=x} = P_Y$).
(ii) Pour tout $y \in F$ tel que $P(Y = y) > 0$, la loi conditionnelle de X sachant $Y = y$ est celle de X (i.e. $P_{X|Y=y} = P_X$).
(iii) Pour tout $(x, y) \in E \times F$, $P\big((X, Y) = (x, y)\big) = P(X = x)\ P(Y = y)$.

Démonstration
Pour la démonstration, on va montrer (i)⇔(iii), on en déduira (ii)⇔(iii) par symétrie, en échangeant les rôles de X et de Y.
Supposons donc (i). Si $y \in F$, si $P(X = x) > 0$,

$$\begin{aligned} P\big((X, Y) = (x, y)\big) &= P(X = x \text{ et } Y = y) \\ &= P(X = x)\ P_{X=x}(Y = y) \\ &= P(X = x)\ P_{Y|X=x}(\{y\}) \\ &= P(X = x)\ P_Y(\{y\}) \\ &= P(X = x)\ P(Y = y) \end{aligned}$$

Et, si $P(X = x) = 0$, comme $(X, Y) = (x, y)$ implique $(X = x)$ (ce qui signifie que l'événement $(X, Y) = (x, y)$ est inclus dans l'événement $X = x$), on a nécessairement $P\big((X, Y) = (x, y)\big) = 0 = P(X = x)\ P(Y = y)$ pour tout y. On en déduit bien (iii).

Supposons réciproquement (iii), et soit $x \in E$ tel que $P(X = x) > 0$. On a, pour tout $y \in F$,

$$\begin{aligned} P_{Y|X=x}(\{y\}) &= P_{X=x}(Y = y) \\ &= \frac{P(Y = y \text{ et } X = x)}{P(X = x)} \\ &= P(Y = y) \\ &= P_Y(\{y\}) \end{aligned}$$

et comme les probabilités « atomiques » (i.e. des événements élémentaires) caractérisent la loi, on déduit bien (i).

Définition On dit que X et Y sont indépendantes lorsqu'elles vérifient une des trois propriétés précédentes.

b. Caractérisation

Remarque préliminaire : Cette caractérisation peut très bien être prise comme définition de l'indépendance de deux variables aléatoires.

Proposition X et Y sont indépendantes si et seulement si, pour toute partie A de E et toute partie B de F,

$$P((X, Y) \in A \times B) = P(X \in A)\, P(Y \in B)$$

C'est bien naturel, mais démontrons-le quand même :
Supposons X et Y indépendantes ; on peut supposer A et B finies, quitte à les remplacer par $A \cap X(\Omega)$ et $B \cap Y(\Omega)$ respectivement. On notera donc $A = \{x_1, \dots, x_p\}$ et $B = \{y_1, \dots, y_q\}$. Alors

$$((X, Y) \in A \times B) = \bigcup_{(i,j) \in [\![1,p]\!] \times [\![1,q]\!]} (X = x_i \text{ et } Y = y_j)$$

et la réunion est disjointe, donc

$$\begin{aligned} P((X,Y)\in A\times B) &= \sum_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]} P(X=x_i \text{ et } Y=y_j) \\ &= \sum_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]} P(X=x_i)P(Y=y_j) \\ &= \left(\sum_{i=1}^{p} P(X=x_i)\right)\left(\sum_{j=1}^{q} P(Y=y_j)\right) \\ &= P(X\in A)\ P(Y\in B) \end{aligned}$$

La réciproque est assez simple en considérant le cas particulier où A et B sont des singletons.

VII.4 Images de v.a. indépendantes par des fonctions

Proposition Si X et Y sont indépendantes, $f(X)$ et $g(Y)$ le sont (avec des notations "évidentes").

Démonstration : Si a et b sont dans l'ensemble d'arrivée de f et g respectivement, on peut écrire

$$\begin{aligned} P\big((f(X),g(Y))=(a,b)\big) &= P\left((X,Y)\in f^{-1}(\{a\})\times g^{-1}(\{b\})\right) \\ &= P\left(X\in f^{-1}(\{a\})\right)\ P\left(Y\in g^{-1}(\{b\})\right) \\ &= P\big(f(X)=a\big)\ P\big(g(Y)=b\big) \end{aligned}$$

ce qui conclut.

VII.5 Variance, covariance

a. Covariance

Définition Si X et Y sont deux variables aléatoires réelles sur un espace probabilisé fini, on définit leur covariance :

$$\mathrm{Cov}(X,Y) = \mathrm{E}\left((X-\mathrm{E}(X))\,(Y-\mathrm{E}(Y))\right)$$

On calcule :

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}\big(XY - \mathrm{E}(X)Y - \mathrm{E}(Y)X + \mathrm{E}(X)\mathrm{E}(Y)\big)\\ &= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) - \mathrm{E}(Y)\mathrm{E}(X) + \mathrm{E}(X)\mathrm{E}(Y)\\ &= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\end{aligned}$$

On peut donc énoncer :

Formule

$$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$$

b. Covariance de variables indépendantes

Proposition Si X et Y sont des variables aléatoires réelles indépendantes,

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$$

et donc

$$\mathrm{Cov}(X,Y) = 0$$

Démonstration

X et Y sont définies sur un espace probabilisé fini Ω, et on note $X(\Omega) = \{x_1, \dots, x_p\}$ et $Y(\Omega) = \{y_1, \dots, y_q\}$. Alors

$$\begin{aligned}\mathrm{E}(XY) &= \sum_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]} x_i y_j P\big((X,Y) = (x_i, y_j)\big)\\ &= \sum_{(i,j)\in[\![1,p]\!]\times[\![1,q]\!]} x_i y_j P(X = x_i)\ P(Y = y_j)\\ &= \sum_{i=1}^{p}\left(x_i P(X = x_i) \underbrace{\sum_{j=1}^{q} y_j P(Y = y_j)}_{\mathrm{E}(Y)}\right)\\ &= E(Y)E(X)\end{aligned}$$

VII.6 Indépendance de n variables, $n \geq 2$

Définition - Proposition Soit (Ω, P) un espace probabilisé fini, n un entier ≥ 2, $X_1, \ldots, X_n$ des variables aléatoires définies sur cet espace, à valeurs dans des ensembles $E_1, \ldots, E_n$ respectivement. Les propositions suivantes sont équivalentes entre elles :

(i) Pour toutes parties A_1 de E_1,...,A_n de E_n, les événements $(X_1 \in A_1)$,...,$(X_n \in A_n)$ sont mutuellement indépendants.

(ii) Pour toutes parties A_1 de E_1,...,A_n de E_n,

$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \ldots P(X_n \in A_n)$.

(iii) Pour tout $(x_1, \ldots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega)$,

$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \ldots P(X_n = x_n)$

Démonstration Il est judicieux de revoir la définition de l'indépendance mutuelle de n événements avant d'aborder cette démonstration.
(i)⇒ **(ii)** est simple.
(ii)⇒ **(i)** : on doit montrer que, si I est une partie finie de $[\![1, n]\!]$,

$$P\left(\bigcap_{j \in I} (X_j \in A_j)\right) = \prod_{j \in I} P(X_j \in A_j)$$

et apparemment, **(ii)** ne donne cela que pour $J = [\![1, n]\!]$. Mais il suffit de définir $A_j = E_j$ si $j \notin I$ pour pouvoir appliquer **(ii)** (en effet, $(X_j \in E_j) = \Omega$ ne modifie pas l'intersection, et $P(X_j \in E_j) = 1$ ne modifie pas le produit).
(ii)⇒ **(iii)** est simple (on prend les A_i égaux à des singletons).
(iii) ⇒ **(ii)** : On peut supposer les A_i finis (quitte à remplacer A_i par $A_i \cap X_i(\Omega)$). Chaque A_i est réunion finie disjointe de singletons, l'implication n'est plus alors qu'une question d'écriture, pas très amusante, mais signalons donc que toutes les réunions écrites ci-dessous sont finies et disjointes, ce qui autorise à utiliser l'additivité :

$$\begin{aligned}
P(X_1 \in A_1, \dots, X_n \in A_n) &= P\left(\bigcup_{(x_1,\dots,x_n)\in A_1\times\dots\times A_n} \bigcap_{i=1}^{n} (X_i = x_i)\right) \\
&= \sum_{(x_1,\dots,x_n)\in A_1\times\dots\times A_n} P\left(\bigcap_{i=1}^{n} (X_i = x_i)\right) \\
&= \sum_{(x_1,\dots,x_n)\in A_1\times\dots\times A_n} \left(\prod_{i=1}^{n} P(X_i = x_i)\right) \\
&= \prod_{i=1}^{n} \left(\sum_{x_i\in A_i} P(X_i = x_i)\right) \\
&= \prod_{i=1}^{n} P(X_i \in A_i)
\end{aligned}$$

VII.7 Variance d'une somme

L'espérance d'une somme de variables aléatoires est la somme des espérances de ces variables. Mais la variance de la somme n'est pas la somme des variances !

a. Variance d'une somme de deux variables

On considère deux variables aléatoires X et Y sur le même espace probabilisé (Ω, P) ;

$$\begin{aligned}
\mathrm{V}(X+Y) &= \mathrm{E}\left((X+Y)^2\right) - (\mathrm{E}(X+Y))^2 \\
&= \mathrm{E}(X^2) + \mathrm{E}(X^2) + 2\mathrm{E}(XY) - (\mathrm{E}(X))^2 - (\mathrm{E}(Y))^2 - 2\mathrm{E}(X)\mathrm{E}(Y) \\
&= \mathrm{V}(X) + \mathrm{V}(Y) + 2\mathrm{Cov}(X,Y)
\end{aligned}$$

On a donc

Formule

$$\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y) + 2\mathrm{Cov}(X,Y)$$

Evidemment, plutôt qu'un calcul direct, cette formule est une formule euclidienne ; en munissant l'ensemble des variables aléatoires sur (Ω, P) du produit scalaire

$$(X|Y) = E(XY)$$

et en notant $\|.\|$ la norme euclidienne associée, alors

$$\mathrm{V}(X) = \|X - \mathrm{E}(X)\|^2$$

Et donc (développement d'un carré scalaire) :

$$\begin{aligned}\mathrm{V}(X+Y) &= \|X+Y-\mathrm{E}(X+Y)\|^2 \\ &= \|X-\mathrm{E}(X)+Y-\mathrm{E}(Y)\|^2 \\ &= \|X-\mathrm{E}(X)\|^2 + \|Y-\mathrm{E}(Y)\|^2 + 2\,(X-\mathrm{E}(X)\,|\,Y-\mathrm{E}(Y)) \\ &= \mathrm{V}(X)+\mathrm{V}(Y)+2\mathrm{Cov}(X,Y)\end{aligned}$$

On voit là encore tout l'intérêt du point de vue euclidien.

b. Cas d'indépendance

Proposition Si X et Y sont indépendantes,

$$\mathrm{V}(X+Y) = \mathrm{V}(X)+\mathrm{V}(Y)$$

ou, si l'on préfère, la variance d'une somme de deux variables aléatoires est égale à la somme des variances de ces variables.

c. Variance d'une somme de n variables aléatoires

On considère n variables aléatoires $X_1, \ldots, X_n$ sur le même espace probabilisé (Ω, P) ;

Formule

$$\mathrm{V}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \mathrm{V}(X_i) \;+\; 2 \sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j)$$

Il s'agit là encore d'une formule de développement d'un carré scalaire (une démonstration par récurrence est donc simple).

d. Cas d'indépendance

Proposition La variance d'une somme de variables aléatoires deux à deux indépendantes est égale à la somme des variances de ces variables aléatoires.

(Il n'est pas nécessaire de supposer ces variables mutuellement indépendantes).

VII.8 Application à la loi binomiale

Proposition Si $X_1, \ldots, X_n$ sont mutuellement indépendantes et suivent une loi de Bernoulli $\mathscr{B}(1,p) = \mathscr{B}(p)$, la somme $X_1 + \ldots + X_n$ suit une loi binomiale $\mathscr{B}(n,p)$

(c'est bien sûr le même p pour toutes les X_i)

Démonstration : Chaque X_i peut prendre deux valeurs : 0 ou 1. Donc $Y = X_1 + \ldots + X_n$ est à valeurs dans $[\![0,n]\!]$. Soit P_k l'ensemble des parties à k éléments de $[\![0,n]\!]$. Une somme de termes égaux à 0 ou à 1 étant le nombre de ces termes égaux à 1, on a

$$\{Y = k\} = \bigcup_{I \in P_k} A_I$$

où

$$A_I = \bigcap_{i \in I} \{X_i = 1\} \cap \bigcap_{i \in [\![1,n]\!] \setminus I} \{X_i = 0\}$$

Les A_I sont deux à deux disjoints ; par indépendance mutuelle,

$$P(A_I) = p^k (1-p)^{n-k}$$

et donc, le cardinal de P_k étant bien connu et la réunion étant disjointe :

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Le résultat est démontré. La variance d'une variable de loi de Bernoulli $\mathscr{B}(p)$ étant $pq = p(1-p)$, on retrouve très simplement (et sans formule de sommation pénible) que la variance d'une loi binomiale $\mathscr{B}(n,p)$ vaut npq.

Table des matières

I Probabilités sur un ensemble fini **2**
I.1 Univers, événements, probabilité 2
I.2 Quelques univers 4
a. Poker 4
b. Bridge 4
c. Tirages répétés 4
d. Des objets dans des boîtes 4
e. Des objets indiscernables dans des boîtes 5
f. Des objets indiscernables dans des boîtes indiscernables . 5
I.3 Première propriété : additivité 5
I.4 Probabilité uniforme 6
I.5 Autres propriétés 7

II Conditionnement, indépendance **11**
II.1 Conditionnement 11
II.2 Probabilités composées, probabilités totales, probabilités des causes 13
II.3 Indépendance 18

III Variables aléatoires **21**
III.1 Définition ; loi d'une variable aléatoire 21
a. Un cadre et une difficulté : le jet de deux dés 21
b. Une variable aléatoire 22
c. Définitions 24
d. Variable en analyse et en probabilités 25
e. Exemple-exercice : une urne de Polya 25
III.2 Espérance d'une variable aléatoire réelle 26
a. Définition 26
b. Propriétés de l'espérance 29
III.3 Variance et écart-type 30
a. Définitions 30
b. Deux formules 31
c. Variable centrée réduite 33

IV Lois usuelles **34**
IV.1 Loi uniforme . . . 34
IV.2 Loi de Bernoulli . . . 34
a. Définition . . . 34
b. Variable de Bernoulli et fonction indicatrice d'un événement 35
IV.3 Loi binomiale . . . 35
a. Définition . . . 35
b. Espérance, variance . . . 36

V Inégalités classiques **39**
V.1 Inégalité de Markov . . . 39
V.2 Inégalité de Bienaymé-Tchebychev . . . 41

VI Image d'une variable aléatoire par une fonction **42**
VI.1 Définition-Notation . . . 42
VI.2 Loi . . . 42
VI.3 Espérance (formule de transfert) . . . 43

VII Couple de variables aléatoires **45**
VII.1 Loi conjointe, lois marginales . . . 45
a. Exemple . . . 45
b. Définitions . . . 46
VII.2 Loi conditionnelle . . . 47
VII.3 Couple de variables aléatoires indépendantes . . . 48
a. Définition . . . 48
b. Caractérisation . . . 49
VII.4 Images de v.a. indépendantes par des fonctions . . . 50
VII.5 Variance, covariance . . . 50
a. Covariance . . . 50
b. Covariance de variables indépendantes . . . 51
VII.6 Indépendance de n variables, $n \geq 2$ . . . 52
VII.7 Variance d'une somme . . . 53
a. Variance d'une somme de deux variables . . . 53
b. Cas d'indépendance . . . 54
c. Variance d'une somme de n variables aléatoires . . . 54

d. Cas d'indépendance 54
VII.8 Application à la loi binomiale 55